

**SYSTEM AND METHOD FOR ANNOTATING MULTI-MODAL**  
**CHARACTERISTICS IN MULTIMEDIA DOCUMENTS**

**Field of the Invention**

The present invention relates to the computer processing of multimedia files.

5 More specifically, the present invention relates to the manual annotation of multi-modal events, objects, scenes, and audio occurring in multimedia files.

**Background of the Invention**

Multimedia content is becoming more common both on the World Wide Web and local computers. As the corpus of multimedia content increases, the indexing of  
10 features within the content becomes more and more important. Observing both audio and video simultaneously and annotating that observation results in a higher confidence level.

Existing multimedia tools provide capabilities to annotate either audio or video separately, but not as a whole. (An example of a video-only annotation tool is  
15 the IBM MPEG7 Annotation Tool, inventors J. Smith et al., available through [\[http://\]www.alphaworks.ibm.com/tech/videoannex](http://www.alphaworks.ibm.com/tech/videoannex). Other conventional arrangements are described in: Park et al, "iMEDIA-CAT: Intelligent Media Content Annotation Tool", Proc. International Conference on Inductive Modeling (ICIM)

2001, South Korea, November, 2001; and Minka et al., "Interactive Learning using a Society of Models," Pattern Recognition, Vol. 30, pp. 565, 1997, TR #349.

It has long been recognized that annotating video or audio features in isolation results in a less confidence of the identification of the features.

5 In view of the foregoing, a need has been recognized in connection with providing improved systems and methods for observing and annotating multi-modal events, objects, scenes, and audio occurring in multimedia files.

### **Summary of the Invention**

In accordance with at least one presently preferred embodiment of the present  
10 invention, there are broadly contemplated multimedia annotation systems and methods that permit users to observe solely video, video with audio, solely audio, or audio with video and to annotate what has been observed.

In one embodiment, there is provided a computer system which has one or more multimedia files that are stored in a working memory. The multi-modal  
15 annotation process displays a user selected multimedia file, permits the selection of a mode or modes to observe the file content, annotates the observations; and saves the annotations in a working memory (such as a MPEG-7 XML file).

In summary, one aspect of the invention provides an apparatus for managing multimedia content, the apparatus comprising: an arrangement for supplying

multimedia content; an input interface for permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and an arrangement for annotating observations of a selected mode.

5           A further aspect of the invention provides a method of managing multimedia content, the method comprising the steps of: supplying multimedia content; permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and annotating observations of a selected mode.

10           Furthermore, an additional aspect of the invention provides a program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for managing multimedia content, the method comprising the steps of: supplying multimedia content; permitting the selection, for observation, of at least one of the following modes associated with the multimedia  
15 content: an audio portion that includes video; and a video portion that includes audio; and annotating observations of a selected mode.

For a better understanding of the present invention, together with other and further features and advantages thereof, reference is made to the following description, taken in conjunction with the accompanying drawings, and the scope of  
20 the invention will be pointed out in the appended claims.

**Brief Description of the Drawings**

Figure 1 is a block diagram depicting a multi-modal annotation system.

Figure 2 is an illustration of a system annotating video scenes, objects, and events.

5        Figure 3 is an illustration of a system annotating audio with video.

Figure 4 is an illustration of a system annotating audio without video.

**Description of the Preferred Embodiments**

Figure 1 is a block diagram of one preferred embodiment of a multi-modal annotation system in accordance with the present invention. The multimedia content and previous annotations are stored on the storage medium 100. When a user 130  
10        selects a multimedia file via the annotation tool from the storage medium 100, it is loaded into working memory 110 and portions of it displayed in the annotation tool 120. At any time, the user 130 may also request that previously saved annotations associated with the current multi-modal file be loaded from the storage medium 100  
15        into working memory 110. The user 100 views the multimedia data by making requests through the annotation tool 120. The user 130 then annotates his observations and the annotation tool 120 saves these annotations in working memory 110. The user can at anytime request the annotation tool 120 to save the annotation on the storage medium 100.

Figure 2 is an illustration of a system annotating video scenes, objects, and events. (Simultaneous reference should also be made to Fig. 1.) The multimedia data has been loaded from the storage medium 100 into working memory 110. A video tab 290 has been selected. The multimedia video has been segmented using scene  
5 changed detection into shots. A shot list window 200 displays a portion of the shots in the multimedia. Here, the user 130 has selected a shot 210 which is highlighted in the shot list window 200. A key frame 220, which is a representative shot in the frames of a shot, is preferably displayed. In addition, the frames of the shot maybe viewed in the video window 230 using play controls 240. The video can be viewed  
10 with or without audio depending upon the selection of a mute button 250. The user 130 may select annotations for this shot by clicking the boxes in events 260, static scenes 270, or key objects 280 lists of boxes. Any significant observations which are not contained in the check boxes can be noted in a keywords text box 300.

Figure 3 is an illustration of the system annotating audio with video.  
15 (Simultaneous reference should also be made to Fig. 1.) The multimedia data has been loaded from the storage medium 100 into working memory 110. The audio with video tab 370 has been selected. The multimedia video has been segmented using scene change detection into shots. The shot list window 200 displays a portion of the shots in the multimedia. The shot 210 associated with the current audio position is  
20 highlighted in the shot list window 200. The audio data is displayed in the window 390. A segment of audio 340 has been delimited for annotation; that is, the limits or

bounds of the audio has been fixed for subsequent annotation. The video associated with the audio is shown in 230. As the user 130 uses the play controls 360, the audio data display 390 is updated to display the current audio data and the video window 230 changes to reflect the current video frame. Thus, the user 130 may observe the video and simultaneously hear the audio while making audio annotations. The user 130 preferably uses the buttons 350 to delimit audio segments. Check boxes corresponding to the foreground sounds (320) (the most prominent sounds in the segment) and background sounds (330) (sounds which are present but are secondary to other sounds) may be checked to indicate sounds heard within the audio segment 340. Any significant observations which are not contained in the check boxes can be noted in keywords text box 300.

Figure 4 is an illustration of the system annotating audio without video. (Simultaneous reference should be made to Fig. 1.) The multimedia data has been loaded from the storage medium 100 into working memory 110. Audio-without-video tab 400 has been selected. The audio data is displayed in the window 390. A segment of audio 340 has been delimited for annotation. As the user 130 uses the play controls 360, the audio data display 390 is updated to display the current audio data. Thus, the user 130 may only hear the audio while making audio annotations. The user 130 uses the buttons 350 to delimit audio segments. The check boxes for foreground sounds 320 and background sounds 330 may be checked to indicate sounds heard within the

audio segment 340. Any significant observations which are not contained in the check boxes can be noted in the keywords text box 300.

It is to be understood that the present invention, in accordance with at least one presently preferred embodiment, includes an arrangement for supplying multimedia  
5 content, an input interface for permitting the selection, for observation, of a mode associated with the multimedia content, and an arrangement for annotating observations of a selected mode. Together, these elements may be implemented on at least one general-purpose computer running suitable software programs. These may also be implemented on at least one Integrated Circuit or part of at least one Integrated  
10 Circuit. Thus, it is to be understood that the invention may be implemented in hardware, software, or a combination of both.

If not otherwise stated herein, it is to be assumed that all patents, patent applications, patent publications and other publications (including web-based publications) mentioned and cited herein are hereby fully incorporated by reference  
15 herein as if set forth in their entirety herein.

Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be affected therein by one skilled in the art without  
20 departing from the scope or spirit of the invention.